SPRINGER BRIEFS IN COMPLEXITY

Francesco Corea

Applied Artificial Intelligence: Where AI Can Be Used in Business

Springer

VISIT…

LANZAROTE
Caliente.COM

SpringerBriefs in Complexity

Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems—cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatialor functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction,freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The three major book publication platforms of the Springer Complexity programare the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations, and the "SpringerBriefs in Complexity" which are concise and topical working reports, case-studies, surveys, essays and lecture notes of relevance to the field. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

More information about this series at http://www.springer.com/series/8907

Francesco Corea

Applied Artificial Intelligence: Where AI Can Be Used in Business

Springer

Francesco Corea
Rome
Italy

ISSN 2191-5326 ISSN 2191-5334 (electronic)
SpringerBriefs in Complexity
ISBN 978-3-319-77251-6 ISBN 978-3-319-77252-3 (eBook)
https://doi.org/10.1007/978-3-319-77252-3

Library of Congress Control Number: 2018934403

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG
part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Introduction

As the title on the front page would suggest, this is a book that basically deals with Artificial Intelligence (AI) and its several applications. It does not simply tackle all the specific cases or sectors where AI can be used, but rather select a few verticals which are very representative of the impact and progresses AI has made, as well as discuss a few operational and strategic issues that any organization may incur in.

In that sense, this book is not an organic text that should be read from the first page onwards, but rather a collection of articles that can be read at will (or at need). The structure of the chapter is very similar, so I hope the reader won't find difficulties in establishing comparisons or understanding the differences between specific problems AI is being used for.

The first chapter of the book is going to show the potential and the achievements of new AI techniques in the speech recognition domain, and it will discuss what a bot is and how this sector would develop, as well as how to classify the players of this space.

The second and third chapters tackle instead verticals that are historically data-intensive but not data-driven, i.e., the financial sector and the insurance one.

The fourth chapter is probably the most innovative because it looks at AI and its intersection with another exponential technology, namely, the blockchain. It analyzes how AI will change the blockchain but also (and more importantly) how the blockchain will foster the development of a decentralized AI.

The last two chapters are instead more operative because they concern new figures to be hired regardless of the organization or the sector, and ethical and moral issues related to the creation and implementation of new types of algorithms.

Chapter 1
AI and Speech Recognition

Abstract This chapter talks about the impact of AI on speech recognition and conversational interfaces. It will touch upon bots classification, main models used and developed in speech recognition as well as a market classification where major start-up players are identified.

1.1 Conversational Interfaces

Conversational User Interfaces (CUI) are at the heart of the current wave of AI development. Although many applications and products out there are simply "*Mechanical Turks*"—which means machines that pretend to be automatized while a hidden person is actually doing all the work—there have been many interesting advancements in speech recognition from the symbolic or statistical learning approaches.

In particular, deep learning is drastically augmenting the abilities of the bots with respect to traditional NLP (i.e., bag-of-words clustering, TF-IDF, etc.) and is creating the concept of "*conversation-as-a-platform*", which is disrupting the apps market.

Our smartphone currently represents the most expensive area to be purchased per squared centimeter (even more expensive than the square meters price of houses in Beverly Hills), and it is not hard to envision that having a bot as unique interfaces will make this area worth almost zero.

None of these would be possible though without heavily investing in speech recognition research. Deep Reinforcement Learning (DFL) has been the boss in town for the past few years and it has been fed by human feedbacks. However, I personally believe that soon we will move toward a B2B (bot-to-bot) training for a very simple reason: ***the reward structure***. Humans spend time training their bots if they are enough compensated for their effort.

This is not a new concept, and it is something Li Deng (Microsoft) and his group are really aware of. He actually provides a great threefold classification of AI bots:

F. Corea, *Applied Artificial Intelligence: Where AI Can Be Used In Business*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-77252-3_1

- Bots that look around for information;
- Bots that look around for information to complete a specific task; and
- Bots with social abilities and tasks (which he names *social bots* or *chatbots*).

For the first two, the reward structure is indeed pretty easy to be defined, while the third one is more complex, which makes it more difficult to be approached nowadays.

When this third class will be fully implemented, though, we would find ourselves living in a world where machines communicate among themselves and with humans in the same way. In this world, the ***bot-to-bot*** business model will be something ordinary and it is going to be populated by two types of bots: ***master bots*** and ***follower bots***.

I believe that research in speech recognition adds up, as well as the technology stacks in this specific space. This would result in some players creating "universal" bots (master bots) which everyone else will use as gateways for their (peripheral) interfaces and applications. The good thing of this centralized (and almost monopolistic) scenario is, however, that in spite of the two-level degree of complexity, we won't have the black box issue affecting the deep learning movement today because bots (either master or follower) **will communicate between themselves in plain English rather than in any programming language**.

1.2 The Challenges Toward Master Bots

Traditionally, we can think of deep learning models for speech recognition as either ***retrieval-based*** models or ***generative models***. The first class of models uses heuristics to draw answers from predefined responses given some inputs and context, while the latter generates new responses from scratch each time.

The state of the art of speech recognition today has raised a lot since 2012, with deep-q networks (DQNs), deep belief networks (DBN), long short-term memory RNN, Gated Recurrent Unit (GRU), sequence-to-sequence learning (Sutskever et al. 2014), and tensor product representations (for a great overview on speech recognition, look at Deng and Li 2013).

So, if DFL breakthroughs were able to improve our understanding of the *machine cognition*, what is preventing us from realizing the perfect social bots? Well, there are at least a couple of things I can think of.

First of all, **machine translation** is still in its infancy. Google has recently created a "Neural Machine Translation", a relevant leap ahead in the field, with the new version even enabling ***zero-short translation*** (in languages which they were not trained for).

Second, speech recognition is still mainly a supervised process. We might need to put further effort into unsupervised learning, and eventually even better integrate the symbolic and neural representations.

Furthermore, there are many nuances of human speech recognition which we are not able to fully embed into a machine yet. MetaMind is doing a great work in the space and it recently introduced **Joint Many-Tasks** (JMT) and the **Dynamic**

Coattention Network (DCN), respectively, an **end-to-end trainable model** which allows collaboration between different layers and a network that reads through documents having *an internal representation of the documents conditioned on the question that it is trying to answer*.

Finally, the automatic speech recognition (ASR) engines created so far were either lacking *personality* or completely *missing the spatiotemporal context*. These are two essential aspects for a general CUI, and only a few works have been tried up to date (Yao et al. 2015; Li et al. 2016).

1.3 How Is the Market Distributed?

This was not originally intended as part of this chapter, but I found useful to go quickly through main players in the space in order to understand the importance of speech recognition in business contexts.

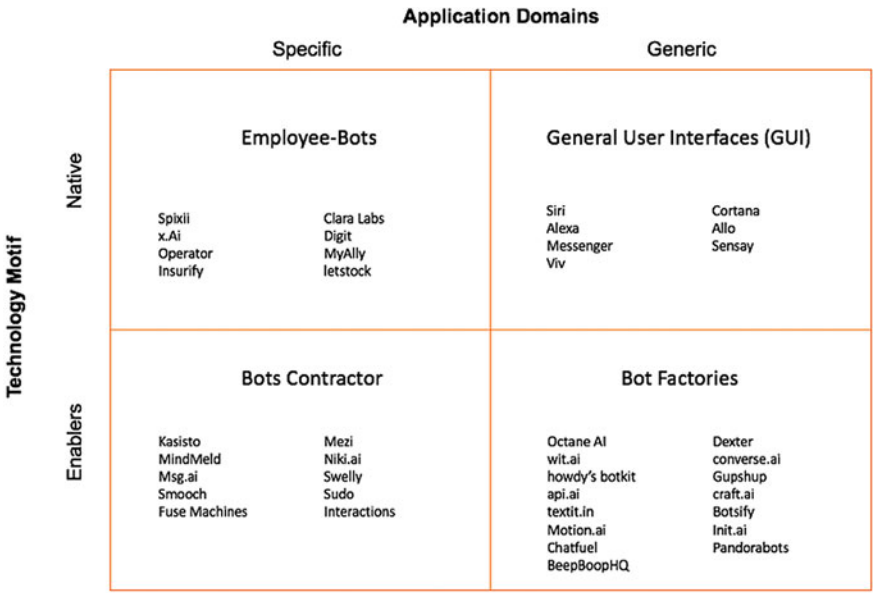
The history of bots goes back to Eliza (1966, the first bot ever), Parry (1968) to eventually ALICE and Clever in the 90s and Microsoft Xiaoice more recently, but it evolved a lot over the last 2–3 years.

I like to think about this market according to this two-by-two matrix. You can indeed classify bots as native or enablers, designed for either specific or generic applications. The edges of this classification are only roughed out and you might actually have companies operating at the intersection between two of these quadrants (Fig. 1.1):

Following this classification, we can identify four different types of start-ups:

- ***Employee Bots***: these are bots that have been created within a specific industry or areas of application. They are stand-alone frameworks that do not necessitate extra training but are ready to plug and play;
- ***General User Interfaces***: these are native applications that represent the purest aspiration to a general conversational interface;
- ***Bots Contractors***: bots that are "hired" to complete specific purposes, but that were created as generalists. Usually cheaper and less specialized than their employee brothers, live in a sort of symbiotic way with the parent application. It could be useful to think about this class as functional bots rather than industry experts (first class);
- ***Bots Factories***: start-ups that facilitate the creation of your own bot.

A few (non-exhaustive) examples of companies operating in each group have been provided, but it is clear how this market is becoming crowded and really profitable.



Application Domains

Technology Motif	Specific	Generic
Native	**Employee-Bots** Spixii x.Ai Operator Insurify Clara Labs Digit MyAlly letstock	**General User Interfaces (GUI)** Siri Alexa Messenger Viv Cortana Allo Sensay
Enablers	**Bots Contractor** Kasisto MindMeld Msg.ai Smooch Fuse Machines Mezi Niki.ai Swelly Sudo Interactions	**Bot Factories** Octane AI wit.ai howdy's botkit api.ai textit.in Motion.ai Chatfuel BeepBoopHQ Dexter converse.ai Gupshup craft.ai Botsify Init.ai Pandorabots

Fig. 1.1 Bots classification matrix

1.4 Final Food For Thoughts

It is an exciting time to be working on deep learning for speech recognition. Not only the research community but the market as well are quickly recognizing the importance of the field as an essential step to the development of an AGI.

The current state of ASR and bots reflect very well the distinction between narrow AI and general intelligence, and I believe we should carefully manage the expectations of both investors and customers. I am also convinced is not a space in which everyone will have a slice of the pie and that a few players will eat most of the market, but it is so quick-moving that is really hard to make predictions on it.

References

Deng, L., & Li, X. (2013). Machine learning paradigms for speech recognition: An overview. *IEEE Transaction on Audio, Speech, and Language Processing, 21*(5).

Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., & Dolan, W. B. (2016). A persona-based neural conversation model. *Association for Computational Linguistics, 1.*

Sutskever, I., Vinyals, O., & Le, Q. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 3104–3112.

Yao, K., Zweig, G., & Peng, B. (2015). Attention with intention for a neural network conversation model. arXiv preprint arXiv:1510.08565

Chapter 2
How AI Is Changing the Insurance Landscape

Abstract This chapter is focusing on insurance and how AI is disrupting the sector. Starting from the conventional insurance process, we will move through specific novelties AI is introducing in the field to understand how insurance is completely changing the way people buy and think of insurance products. Finally, an eight-group classification is proposed for the insurtech ecosystem.

2.1 A Bit of Background

The insurance sector is one of the most old-fashioned and resistant-to-change space, and this is why AI will have a greater impact on that with respect to more receptive industries. The collection of data of new types (i.e., unstructured data such as reports, images, contracts, etc.) and the use of new algorithms are disrupting the sector in several ways.

Traditionally, an insurance company followed this type of process:

- Identifying pool of customers whom might be risk-assessed;
- Targeting those customers and assessing the risk for each class;
- Selling differently priced policies spreading the risks over the pool of customers;
- Try to retain those customers as long as possible offering lower price for longer contracts.

This is a really simplistic representation of the insurance business in the last 50 years, and I am aware that insurance experts might disagree with me in many different ways. There are a couple of further features to be pointed out: first of all, insurance has historically been ***sold not bought***, which means that brokers and agents were essential to tracking new customers and to even retain old ones. In addition, it is an industry which is by definition rich of data because they collected anything they could, but is also one of the less advanced because either many of those data are unstructured or semi-structured, or the model used are quite old and simple.

F. Corea, *Applied Artificial Intelligence: Where AI Can Be Used In Business*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-77252-3_2

Most of those data were easy to obtain because they were required to correctly price the coverage, while additional complimentary data were provided only by good customers who had incentives in providing as much data as possible to get a cheaper policy. Of course, this works the other way for bad customers, and this is a perspective on the phenomenon of "*adverse selection*" (i.e., bad customers are going to ask an insurance because they feel they will need it).

The adverse selection issue is though only one of the intrinsic challenges of the sector: ***strong regulation***, high level of ***fraud*** attempts, and ***complexity*** are other features any incumbents should take care of. It is interesting to notice though that some of those are also specific barriers to entry for startups: they might attract indeed people who normally can get affordable insurance with a bigger competitor (adverse selection) and they usually have the capabilities for breaking down the risk complexity but not to support the funding need for risk coverages (so they need to work with incumbents rather than trying to replace them).

In spite of those problems, in the last decade, we noticed a new trend emerging. Insurances, in the effort of trying to reduce *moral hazard* problems, they started offering *premium discounts to their final customers* in order to get extra information. This occurred either through a questionnaire (asking **directly** the customer for further data in exchange for a lower price) or **indirectly** through devices (healthy devices, black boxes, etc.). The real issue though has been the engagement side of this proposal, because of the opposite nature of information, rewards, and human nature. The rewards offered were indeed either temporary or provided only once and people got lazy very quickly, while the information stream needed to be constant.

The following step has been the introduction of apps to let customers monitor by themselves their own data and behavior, sometimes even given away for free the device itself. Leaving the customer with full power on his data had though an inverse effect, because people did not have the motivation in tracking down their improvements, and they got upset at the same time because they felt they were not getting the most out of that opportunity.

Regardless of the specific innovative way in which insurers engaged customers, the process used in the insurance business did not change much in the past century. ***Expert systems*** and ***knowledge engineering*** dominated the sector setting the rules to be followed in internal workflows, but this is slowly changing with intelligent automation systems. We are actually migrating from rule-based decision systems to statistical learning and eventually machine learning.

2.2 So How Can AI Help the Insurance Industry?

AI is helping (or disrupting, depending on how you see the matter) the sector in different ways. First of all, it can help ***increasing the customer engagement and retention*** problem which has been just mentioned. The abundance of data can be used indeed to refine the customers' segmentation and provide personalized offers based on personal features. It also helps in ***reducing the costs*** through smart automatization or RPA (robotic process automation).

Second, AI is making people ***more aware of the risks as well as habits***, and it is driving them toward better behaviors.

Furthermore, the better pricing and risk assessment that AI is introducing analyzing more granular data will make some people ***uninsurable*** (i.e., too risky to be fairly priced and covered) as well as to turn back some previously uninsurable people into insurable customers again. The governments or central regulatory agencies should then start thinking about a "*pricing/risk threshold*" in which they intervene subsidizing the cost of relevant insurances (e.g., basic health coverage) in order to "*guarantee the uninsurables*".

Finally, it might be useful to think in terms of what an insurable risk is in order to see how AI can help with that.

According to Jin Park (Assistant Professor at IWU), an insurable risk is identifiable through the following five conditions:

- Large number of similar exposure units (mutuality);
- Accidental and unintentional loss (not predictable and independent from the insured customers);
- Determinable and measurable loss;
- Calculable chance of (not catastrophic/systemic) loss;
- Economically feasible premium.

AI is going to affect all those features: with a better and more detailed customer profiling, we won't need indeed to have such a large base of insured units. It will turn some frequent events into accidental (e.g., affecting drivers' behavior it will reduce the basic accidents into rare events) and it will improve our ability to forecast and compute both the probability and magnitude potential losses even in those cases too hard to be managed before. All the previous improvements will make many more premium under budgets, and therefore the conclusion is that AI will "*lower*" the threshold of what we consider nowadays an insurable risk, and it will make then more risks insurable.

2.3 Who Are the Sector Innovators?

There are plenty of start-ups out there working at the intersection of AI and insurance, and it essential to look at least at some of them to understand the future direction of the industry, as well as the kind of improvements AI is having in the insurtech space. An interesting thing to notice is that most of the innovation is happening in the UK rather than other countries, in all the segments proposed below.

Claim Processing: Shift Technology skims the valid claims from the ones that deserve further validations; Tractable instead is trying to automatize experts task for insurances; ControlExpert has a specific focus on car claims; Cognotekt optimizes internal business processes, as well as Snapsheet does; Motionscloud offers instead mobile claim management solutions; and finally RightIndem aims to help insurers to deliver on-premise smoothing the claiming flow.

Virtual Agents and Chatbots: Spixii is an automated insurance agent who helps you buying any insurance coverage you might want; Cognicor is a virtual assistant that offers customer care services; Conversica identifies which leads intend to purchase, while Your. MD is a personal health assistant that analyzes symptoms and produces pieces of advice. MedWhat instead uses EMR (medical records) to assist the patient as it was a virtual doctor, and Babylon gives medical advice taking care of tight budget constraints. Insurify is another personal insurance agent who works as a comparator for car insurances.

What today is called simply chatbot is going to be renamed in a few years **robo-insurer**. There are already few examples of companies toward that goal: Risk Genius is indeed an intelligent comparator which identifies gaps in coverage for the customer and PolicyGenius looks for the best solution that fits customer's needs and characteristics, while Drive Spotter implements real-time video analytics to keep drivers safe(r). More generally, robo-insurers will be a quite wide class of agents who will end up providing different services, all of them with the final goal of helping the clients to undertake risk-mitigating actions and only cover the real (residual) risks.

Customers Engagement: Oscar is probably the most successful insurtech company out there, with the final goal of making insurance simple and accessible to everyone through a great UX. Similar to Oscar is somehow Stride Health, while Brolly is a tool that helps customers in understanding their own needs and facilitates in one place all the insurance coverages in place, in a similar fashion to Knip. Adtelligence instead creates personalized offers and relevant products based on customer's characteristics. Captricity uses machine learning to convert handwritten files into structured data, and this can be used to better understand the final customer. Finally, ValChoice ranks the service of insurers to the benefit of the client.

Telematics: connected cars and telematics is a pretty big area itself, but it would be worthy to point out the work that Greenroad, Vnomics, and Telogis are doing in capturing driving behaviors and habits as well as computing fuel efficiency. Cambridge Mobile Telematics works similarly, although it uses smartphone data and mobile devices habits. Navdy is trying to revolutionizing the UI/UX within vehicles, displaying information in such a way that the driver does not get distracted. Lytx uses vision technology to provide real-time feedbacks to the driver.

Underwriting: AI can be (and actually is) used to spot out hidden correlations to granularly segment customers and risks in a more efficient way. Even though it might in theory possible to identify some algos that could perform better than others (see the work Wipro did for fraud detection), data always come first, at least for the next close future. Many companies operate in the space, as for instance Carpe Data that provides predictive algorithms and data products for property and casualty and life insurances through the analysis of public data (e.g., social media data). Atidot created a machine learning risk management platform, while Tyche uses unstructured data to optimize the underwriting and claims process. Big Cloud Analytics collects data from wearables and formulates health scores for a better risk assessment, while Cape Analytics uses computer vision techniques on geospatial data to improve the level of detail on current houses conditions. Dreamquark creates a more accurate representation of the medical datasets to be used for underwriting purposes by insurances, similarly to FitSense that offers also apps products. Melody Health Insurance provides also low-cost insurances, while Uvamo uses AI to assess the risk of policy applications. A more accurate underwriting can even translate into covering events that are today quite risky (e.g., as MeteoProtect and Praedicat, and are doing for weather risk management).

Finally, on a side, it is worthy to point out to pure technological enablers as Instanda, which offers a management tool to the insurance providers to manage effectively and timely new products launched; Insly, a cloud-based platform for insurance brokers; and finally, SimpleInsurance is instead an e-commerce provider for product insurances.

P2P insurance: Lemonade, Friendsurance, and Guevara are peer-to-peer insurance start-ups focusing respectively on property and casualty insurance the first two, and car insurance the latter one.

Insurchain and Smart Contracts: these are companies in the insurance sector that are driven by *blockchain technology*. Elliptic offers real-time AML for bitcoin specifically, while Everledger is a permanent immutable ledger for diamond certification. Luther Systems is instead a stealth-mode company working on the standardization of smart contracts. Dynamis provides a P2P supplementary unemployment insurance product, while Saldo.mx provides micro-insurance policies on the blockchain. SafeShare covers multiple parties with insurance cover at short notice and for varying periods, and finally, Teambrella is another P2P insurance platform run on the blockchain.

Insurance On-demand: this class of start-ups put in customers' hand the entire insurance buying process. Trov is probably the best example of this new class of players and it allows to ensure things by simply taking a picture of them. Cuvva is quite similar but with a focus on car insurance, Sure and Airsurety on travel policies, and Back me up is another example of on-demand insurance. But this class does not include only the proper on-demand business model, but also insurance start-ups which provide products that vary by location, time, use, or customer. In other words, pay-per-mile business model (Metromile), micro-insurance policies (Neosurance), or eventually Insurance-as-a-service models (Digital Risks).

2.4 Concluding Thoughts

Yan identifies four elements which constitute the insurance profit structure: premium earned and the investment income from one hand, and underwriting cost and claim expenses from the other. AI is and will be able to improve the cost structure, increasing at the same time the competitiveness and enlarging the customer base accessible to insurers, while optimizing internal processes and enhancing the transparency and robustness of the compliance flow.

The greatest challenge I still see in insurance is the ***cultural mindset*** which might prevent insurance to adopt early AI solutions, although this won't probably have a long life given the incredible pressure to innovate the insurance providers are undergoing through.

Chapter 3
How AI Is Transforming Financial Services

Abstract Finance and traditional banking is an industry which is historically data-rich although not so much data-driven. AI is affecting the sector in several ways, ranging from financial wellness to financial security, capital markets, and even money transfer. But above all AI is forcing the financial services players to innovate and to look for alternative solutions to old problems.

3.1 Financial Innovation: Lots of Talk, Little Action?

The financial sector is historically one of the most resistant to change you might think of. It is then inevitable that big banks from one hand and start-ups from the other hand are creating a huge break in the financial industry and I believe this is happening not because of the use of a specific technology but rather because of their intrinsic cultural differences, diverse structural rigidity, and alternative cost-effective business models.

In other words, banks do not innovate either because they are too big to quickly adapt and follow external incentives or because they don't know how (and want to) truly change. This is not simply true in the industry but also in academia, where until the mid-90s there were no relevant contributions to financial innovation at all (Frame and White 2002). In fact, in few survey articles (Cohen and Levin 1989; Cohen 1995) with more than 600 different articles and books quoted, **none of them was related to financial innovation subjects**.

Of course, things changed over the last 5 years, but my opinion is that was really out of necessity rather than a voluntary push-approach from the banking sector.

Financial innovation is, therefore, something which seems to be usually ***imported*** rather than ***internally generated***, and often more characterized by a ***product-innovation*** rather than a ***process one*** (although this might be a controversial opinion, I guess). Given the new technological paradigm (which is tightening the inner strong causal relationship between innovation and growth), it seems natural to won-

F. Corea, *Applied Artificial Intelligence: Where AI Can Be Used In Business*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-77252-3_3

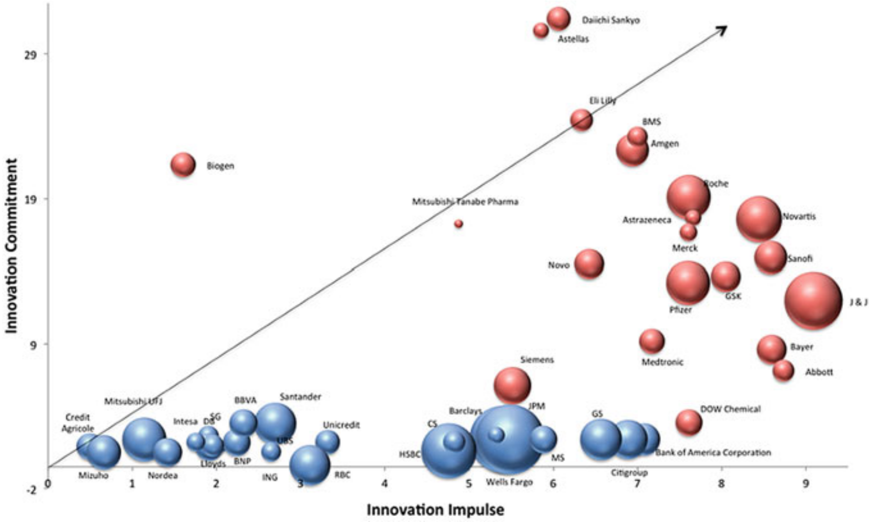


Fig. 3.1 Innovation drivers map (Corea 2015). The biopharma companies feel more the urgency to innovate and are also more committed to that. The graph I built plots 25 major banks (blue) and 25 major pharmaceutical (red) companies based on their Innovation Impulse and Commitment. The Impulse variable has been built using the number of patents a company filed (a proxy for the external pressure to innovate) and the number of recorded shareholders (a proxy for the internal pressure to innovate). The Commitment shows instead the R&D intensity (net sales) while the size of the bubbles the net income for each company. The data points were obtained by Medtrack, Osiris, and Zephyr in 2014

der whether a better innovation model can be therefore imported by a different (and more successful) sector.

I found that there is a very specific and interesting case of a sector which had to "***innovate-to-survive***" rather than "***innovate-to-grow***": the biopharma industry (Baker 2003; Gans and Stern 2003; Fuchs and Krauss 2003; Lichtenthaler 2008) (Fig. 3.1).

The graph I built plots 25 major banks (blue) and 25 major pharmaceutical (red) companies based on their Innovation Impulse and Commitment. The Impulse variable has been built using the number of patents a company filed (a proxy for the external pressure to innovate) and the number of recorded shareholders (a proxy for the internal pressure to innovate). The Commitment shows instead the R&D intensity (net sales) while the size of the bubbles the net income for each company. The data points were obtained by Medtrack, Osiris, and Zephyr in 2014.

3.2 Innovation Transfer: The Biopharma Industry

The biopharma industry is not a single sector but actually two different ones: the **biotech space**, populated by smaller companies that drive the research and exploration phase, and the **pharmaceutical companies**, big giants that through the last century became huge go-to-market and sales enterprises.

Hence, there is **pure (risky) innovation** from one hand and **pure commercialization skills** from the other…Is it something that we have already seen somewhere, didn't we? The biopharma industry and the financial sector suffer indeed from a strong polarized innovation.

What characterizes the industry is that the risky activity lies in the initial development process rather than in the market phase. The problem is not to match customer demand or find a market for your product, but it is actually developing the molecule in the first place. The probability of success is extremely low and the timeline is very long (10–15 years) and the 20-year patents give you only a temporary advantage. More importantly, it looks that only 3 out of 10 of the drugs produced are able to repay the development costs (Meyer 2002) and that most of the companies operate at loss while the top 3% companies alone generate almost 80% of the entire industry profits (Li and Halal 2002). A tough business, isn't it?

The biopharma industry is then no longer simply a ***human-intensive business*** but also a ***capital-demanding one***. Innovation is not ancillary but it is the quintessential driver to survive. And this is also why they had to identify a range of different methods to foster their growth-by-innovation: R&D, competitive collaboration schemes, venture funding, co-venture creation, built-to-buy deals, limited partnership agreements, etc.

It should be clear by now where I am heading to: **the financial industry doesn't strongly feel the need to innovate** as the biopharma sector and it is **not experimenting and pushing to create new models that might spread their innovation risk** and make it profitable.

3.3 Introducing AI, Your Personal Financial Disruptor

By now you might object "All good man, but financial services and biopharma are still sooo different, so why should I import innovation models from a sector which is completely different from mine?". Well, that's the catch: I don't think they are.

And the reason why they are becoming a lot more similar is precisely **Artificial Intelligence**.

AI is creating a strong pressure to innovate for the financial sector and has a **development cycle and characteristics** which are somehow similar to the biopharmaceutical one: it requires a long time to be created, implemented and correctly deployed (with respect to the financial industry standards, of course); it is highly technical and requires highly specialized talents; it is highly uncertain, because you need to experiment a lot before finding something that works; it is expensive, both in terms of time as well as monetary investments (talents, hardware, and data are really expensive); it is risky and the risk lies in the initial development phase, with a very high payout but a high likelihood to fail as well.

But AI is also introducing a completely new speed and degree of trust in the financial industry, which lowers the tolerable mistakes at the same level of the biopharma sector. If your algorithms point out to the wrong product to sell or the wrong book to be recommended, it is not a big deal. If your system misinterprets some signals in the market or while developing a drug though, you end up **losing millions in seconds or even losing human lives**.

It is then not only stretching out issues that intrinsically belong to the financial sector such as **regulation** or **accountability**, but it is also bringing new problems such as **biased data or the lack of transparency** to the picture (specifically in consumer applications).

And last but not least, AI is making the question mark on the "***build versus buy***" matter bigger than even in FS, the same as it was in the biopharma industry back in the 90s and that culminated in the current biotech-pharmaceutical dichotomy (if you are wondering anyway, this choice is all focused around on your ***data capacity, team and project scalability***, and ***uniqueness of the project*** with respect to your competitors—do you have enough data to train an ANI? Can your team/project scale? Is the ANI unique or something your peers are doing or need to do as well?).

This is why I believe AI in financial services to be extremely important—not much for the specific innovation or product it is introducing but rather because ***it is revolutionizing a centuries-old industry innovation flow from the ground***.

3.4 Segmentation of AI in Fintech

AI is using structured and unstructured data in financial services to improve the customer experience and engagement, to detect outliers and anomalies, to increase revenues, reduce costs, find predictability in patterns and increase forecasts reliability…but it is not so in any other industry? We all know this story, right? So what is really peculiar about AI in financial services?

First of all, financial services is an industry full of data. You might expect this data to be concentrated in big financial institutions' hands, but most of them are actually public and thanks to the new **EU payment directive (PSD2) larger datasets are available** to smaller players as well. AI can then be easily developed and applied because the barriers to entry are lower with respect to other sectors.

Second, many of the underlying processes can be relatively **easier to be automatized** while many others can be improved by either brute force computation or speed. And historically is one of the sectors that needed this type of innovation the most, is incredibly competitive and is always looking for some new source of ROI. **Bottom line: the marginal impact of AI is greater than in other sectors**.

Third, the **transfer of wealth across different generations** makes the field really fertile for AI. AI needs (a lot of) innovative data and above all feedback to improve, and millennials are not only happy to use AI as well as providing feedback, but apparently even less concerned about privacy and giving away their data.

There are also, of course, a series of **specific challenges for AI** in financial sector that limit a smooth and rapid implementation: legacy systems that do not talk to each other; data silos; poor data quality control; lack of expertise; lack of management vision; lack of cultural mindset to adopt this technology.

So what is missing now is only having an overview of the AI fintech landscape. There are also plenty of maps and classification of AI fintech start-ups out there, so I am not introducing anything new here but rather simply giving you my personal framework:

- **Financial Wellness**: this category is about making the end-client life better and easier and it includes ***personalized financial services***; ***credit scoring***; automated financial advisors and planners that assist the users in making financial decisions (***robo-advisor, virtual assistants, and chatbots***); smart wallets that coach users differently based on their habits and needs. *Examples include [robo-advisors and conversational interfaces] Kasisto; Trim; Penny; Cleo; Acorns; Fingenius; Wealthfront; SigFig; Betterment; LearnVest; Jemstep; [credit scoring] Aire; TypeScore; CreditVidya; ZestFinance; Applied Data Finance; Wecash*;
- **Blockchain**: I think that, given the importance of this instrument, it deserves a separate category regardless of the specific application is being used for (which may be payments, compliance, trading, etc.). *Examples include: Euklid; Paxos; Ripple; Digital Asset*;
- **Financial Security**: this can be divided into **identification** (payment security and physical identification—**biometrics and KYC**) and **detection** (looking for fraudulent and abnormal financial behavior—**AML and fraud detection**). *Examples include, respectively: EyeVerify; Bionym; FaceFirst; Onfido; and Feedzai; Kount, APEX Analytics*;
- **Money Transfer**: this category includes payments, peer-to-peer lending, and debt collection. *Examples include: TrueAccord; LendUp; Kabbage; LendingClub*;
- **Capital Markets**: this is a big section, and I tend to divide it into five main subsections:

  - **Trading** (either algotrading or trading/exchange platforms). *Examples include: Euclidean; Quantestein; Renaissance Technologies, Walnut Algorithms; EmmaAI; Aidyia; Binatix; Kimerick Technologies; Pit.ai; Sentient Technologies; Tickermachine; Walnut Algorithm; Clone Algo; Algoriz; Alpaca; Portfolio123; Sigopt*;
  - **Do-It-Yourself Funds** (either crowdsource funds or home-trading). *Examples include: Sentifi; Numerai; Quantopian; Quantiacs; QuantConnect; Inovance*;
  - **Markets Intelligence** (information extraction or insights generation). *Examples include: Indico Data Solutions; Acuity Trading; Lucena Research; Dataminr; Alphasense; Kensho Technologies; Aylien; I Know First; Alpha Modus; ArtQuant*;

- **Alternative Data** (most of the alternative data applications are in capital markets rather than broader financial sector so it makes sense to put it here). *Examples include: Cape Analytics; Metabiota; Eagle Alpha*;
- **Risk Management** (this section is more a residual subcategory because most of the time start-ups in this group fall within other groups as well). *Examples include: Ablemarkets; Financial Network Analysis.*

3.5 Conclusions

I am arguing since the beginning of the article that AI is making financial services and biopharma much more alike, and that the FS industry might learn something from how the other industry innovates.

The reality is that the financial industry has also very specific traits and challenges it needs to overcome.

The biggest difference I currently see in that is the effect AI is having on the physical products market: while in almost any sector AI is used with the final goal of creating or improving new products (and this is true also for drug development, for example) in the financial ecosystem is having exactly the opposite effect. **AI is making the industry more digitalized than ever before**. Its final goal will be to create the (frictionless) bank of the future: no branches, no credit cards, no frauds, no menial reporting activities. A bank-as-a-platform with modular components that increases our financial literacy and has no physical products or spaces.

References

Baker, A. (2003). Biotechnology's growth-innovation paradox and the new model for success. *Journal of Commercial Biotechnology, 9*(4), 286–288.

Cohen, W. (1995). Empirical studies of innovative activity. In P. Stoneman (ed.), *Handbook of the economics of innovation and technological change* (Chap. 6, pp. 182–264). Cambridge, Massachusetts: Blackwell.

Cohen, W., & Levin, R. (1989). Empirical studies of innovation and market structure. In R. Schmalensee & R. Willig (eds.), *Handbook of industrial organization* (Vol. 2, Chap. 18, pp. 1059–1107). Amsterdam: North-Holland.

Corea, F. (2015). What finance can learn from biopharma industry: A transfer of innovation models. *Expert Journal of Finance, 3,* 45–53.

Frame, W. S., & White, L. J. (2002). *Empirical studies of financial innovation: Lots of talk, little action?.* In Working Paper, Federal Reserve Bank of Atlanta, N. 2002–2012.

Fuchs, G., & Krauss, G. (2003). Biotechnology in comparative perspective. In G. Fuchs (Ed.), *Biotechnology in comparative perspective* (pp. 1–13). New York: Routledge.

Gans, J., & Stern, S. (2003). *Managing ideas: Commercialization strategies for biotechnology.* Intellectual Property Research Institute of Australia Working Paper No. 01/03, pp. 1–24.

Lichtenthaler, U. (2008). Open innovation in practice: An analysis of strategic approaches to technology transactions. *IEEE Transactions on Engineering Management, 55*(1), 148–157.

Li, J., & Halal, W. E. (2002). Reinventing the biotech manager. *Nature Biotechnology, 20*(Suppl 6), 61–63.

Meyer, F. J. (2002, November 25). Business models that biotech companies employ. *Enterprise development KFBS biotech speakers series*.

Chapter 4
The Convergence of AI and Blockchain

Abstract This is likely the most technical chapter of the book, although it is explained in a comprehensible language. It deals with the intersection between AI and blockchain and how they are affecting each other. A primer on blockchain is provided, as well as a full map of players working with those two exponential technologies.

4.1 Setting the Stage

I have been talking and writing about AI since a while now, so I will not waste any time defining what it is and what is not.

However, I never touched upon blockchain and cryptocurrencies so far, so I will dedicate this first block to describe what it is and how it works.

A blockchain is a **secure distributed immutable database shared by all parties in a distributed network** where transaction data can be recorded (either *on-chain* for basic information or *off-chain* in case of extra attachments) and easily audited.

Put simply (with Bank of England's words), the blockchain is "a *technology that allows people who don't know each other to trust a shared record of events.*"

The data are stored in rigid structures called **blocks**, which are connected to each other in a **chain** through a *hash* (each block also includes a *timestamp* and a link to the previous block via its *hash*). The blocks have a header, which includes metadata, and a content, which includes the real transaction data. Since every block is connected to the previous one, as the number of participants and blocks grow, it is extremely hard to modify any information without having the network consensus.

The network can validate the transaction through different mechanisms, but mainly through either a "*proof-of-work*" or a "*proof-of-stake*". A **proof-of-work** (Nakamoto 2008) asks the participants (called "*miners*") to solve complex mathematical problems in order to add a block, which in turn require a ton of energy and hardware capacity to be decoded. A **proof-of-stake** (Vasin 2014) instead tries to

F. Corea, *Applied Artificial Intelligence: Where AI Can Be Used In Business*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-77252-3_4

solve this energy efficiency issue attributing (roughly) more mining power to participants who own more coins (there are many variations of it and some skepticism around its famous "*nothing at stake*" problem).

Additional mechanisms are the Byzantine fault-tolerant algorithm (Castro and Liskov 2002), the Quorum slicing (Mazieres 2016), as well as variations of the Proof-of-stake (Mingxiao et al. 2017), but we will not get into those now.

The final characteristic that needs to be explained is the category of blockchain based on the different network access permissions, i.e., whether it is free for anyone to view it (**permissionless versus permissioned**) or to participate in the consensus formation (**public versus private**). In the former case, anyone can access and read or write data from the ledger, while in the latter one predetermined participants have the power to join the network (and of course only in the public permissionless case a reward structure for miners has been designed).

It should be clear by now the intrinsic power of this technology, which is not simply a disruptive innovation but rather a *foundational technology* that aims to "*change the scope of intermediation*" (Catalini and Gans 2017). Distributed ledger technologies will indeed reduce both the costs of verification and networking, influencing then the market structure and eventually allowing the creation of new marketplaces. Iansiti and Lakhani (2017) also drew a brilliant parallel between blockchain and TCP/IP in a recent work (which I highly recommend), showing how blockchain is slowly going through the four phases that identify previous foundational technologies such as the TCP/IP, i.e., single use, localized use, substitution, and transformation. As they explained, the "*novelty*" of such a technology makes it harder for people to understand the solution domain, while its "*complexity*" requires a larger institutional change to foster an easy adoption.

However, it is also true that the blockchain is shifting the traditional business models distributing value in an opposite way with respect to previous stacks: if it made more sense to invest in applications rather than protocol technologies 15 years ago, in a blockchain world, the value is concentrated in the shared protocol layer and only marginally at the application level (see the "***Fat Protocol***" theory by Joel Monegro).

To conclude this introductory section, I will just mention on the fly the possibility for the blockchain to not simply allow for transactions but also the possibility to create **(smart) contracts** that are triggered by specific events and threshold and that are traceable and auditable without effort.

4.2 Bonus Paragraph: Initial Coin Offerings (ICOs)

A big hype is nowadays surrounding this new phenomenon of the *Initial Coin Offerings (ICOs)*. Even if many people are pouring money into that because of its resemblance to the most common (and valuable), Initial Public Offerings (IPOs), an ICO is nothing more than a **token sale**, where a token is the *smallest functional unit of a specific network* (or application).

ICO's experts (if any) will forgive my approximate definition, but an ICO is a hybrid concept that has elements of a ***shares allocation***, a ***pre-sales/crowdfunding campaign***, and a ***currency*** with a limited power and application's domain.

It is definitely an interesting innovation that introduces new unregulated ways to raise capitals, but it also poses several issues to an unprepared community. I am happy to receive feedback on this, but I would distil the key points of an ICO evaluation in what follows:

- A token has an additional utility with respect to the exchange of value and companies selling token with the **only goal of raising capital are sending a bad signal** to the market. Tokens are needed to create a users' base and to incentivize stakeholders to participate in the ecosystem at the earliest stage. **A good white paper is not enough**;
- Be wary of token sales that are **uncapped**;
- Be wary of token sales that have **no time limit**;
- Be wary of token sales that do not clearly state the (present and future) **number** as well as the **value of the token** (it could sound absurd, but you may be surprised of how non-transparent an ICO can look like).

4.3 How AI Can Change Blockchain

Although extremely powerful, a blockchain has its own limitations as well. Some of them are technology-related while others come from the old-minded culture inherited from the financial services sector, but all of them can be affected by AI in a way or another:

- **Energy Consumption**: *mining* is an incredibly hard task that requires a ton of energy (and then money) to be completed (O'Dwyer and Malone 2014). AI has already proven to be very efficient in optimizing energy consumption, so I believe similar results can be achieved for the blockchain as well. This would probably also result in lower investments in mining hardware;
- **Scalability**: the blockchain is growing at a steady pace of 1 MB every 10 min and it already adds up to 85 GB. Nakamoto (2008) first mentioned "*blockchain pruning*" (i.e., deleting unnecessary data about fully spent transactions in order to not hold the entire blockchain on a single laptop) as a possible solution but AI can introduce new decentralized learning systems such as **federated learning**, for example, or new data sharing techniques to make the system more efficient;
- **Security**: even if the blockchain is almost impossible to hack, its further layers and applications are not so secure (e.g., the DAO, Mt Gox, Bitfinex, etc.). The incredible progress made by machine learning in the last 2 years makes AI a fantastic ally for the blockchain to guarantee a secure applications deployment, especially given the fixed structure of the system;
- **Privacy**: the privacy issue of owning personal data raises regulatory and strategic concerns for competitive advantages (Unicredit 2016). ***Homomorphic encryption***

(performing operations directly on encrypted data), the **Enigma project** (Zyskind et al. 2015) or the **Zerocash project** (Sasson et al. 2014), are definitely potential solutions, but I see this problem as closely connected to the previous two, i.e., scalability and security, and I think they will go *pari passu*;

- **Efficiency**: Deloitte (2016) estimated the total running costs associated with validating and sharing transactions on the blockchain to be as much as $600 million a year. An intelligent system might be eventually able to compute on the fly the likelihood for specific nodes to be the first performing a certain task, giving the possibility to other miners to shut down their efforts for that specific transaction and cut down the total costs. Furthermore, even if some structural constraints are present, a better efficiency and a lower energy consumption may reduce the ***network latency*** allowing then faster transactions;
- **Hardware**: miners (and not necessarily companies but also individuals) poured an incredible amount of money into specialized hardware components. Since energy consumption has always been a key issue, many solutions have been proposed and much more will be introduced in the future. As soon as the system becomes more efficient, some piece of hardware might be converted (sometimes partially) for neural nets use (the mining colossus Bitmain is doing exactly this);
- **Lack of Talent**: this is leap of faith, but in the same way we are trying to automate data science itself (unsuccessfully, to my current knowledge), I don't see why we couldn't create virtual agents that can create new ledgers themselves (and even interact on it and maintain it);
- **Data Gates**: in a future where all our data will be available on a blockchain and companies will be able to directly buy them from us, we will need help to grant access, track data usage, and generally make sense of what happens to our personal information at a computer speed. This is a job for (intelligent) machines.

4.4 How Blockchain Can Change AI

In the previous section, we quickly touched upon the effects that AI might eventually have on the blockchain. Now instead, we will make the opposite exercise understanding what impact can the blockchain have on the development of machine learning systems. More in detail, blockchain could:

- **Help AI Explaining Itself (and Making Us Believe It)**: the AI black-box suffers from an explainability problem. Having a clear audit trail can not only improve the **trustworthiness** of the data as well as of the models but also provide a clear route to trace back the machine decision process;
- **Increase AI Effectiveness**: a secure data sharing means more data (and more training data), and then better models, better actions, better results, and better new data. Network effect is all that matter at the end of the day;
- **Lower the Market Barriers to Entry**: let's go step by step. Blockchain technologies can secure your data. So why won't you store all your data privately and

maybe sell it? Well, you probably will. So, first of all, blockchain will foster the creation of **cleaner and more organized personal data**. Second, it will allow the emergence of **new marketplaces**: a ***data marketplace*** (low-hanging fruit); a ***models marketplace*** (much more interesting); and finally even an ***AI marketplace*** (see what Ben Goertzel is trying to do with SingularityNET). Hence, easy data sharing and new marketplaces, jointly with blockchain data verification, will provide a more fluid integration that lowers the barrier to entry for smaller players and shrinks the competitive advantage of tech giants. In the effort of lowering the barriers to entry, we are then actually solving two problems, i.e., providing a ***wider data access*** and a more efficient ***data monetization mechanism***;

- **Increase Artificial Trust**: as soon as part of our tasks will be managed by autonomous virtual agents, having a clear audit trail will help **bots to trust each other** (and us to trust them). It will also eventually increase every **machine-to-machine interaction** (Outlier Ventures 2017) and transaction providing a secure way to share data and coordinate decisions, as well as a robust mechanism to reach a quorum (extremely relevant for swarm robotics and multiple agents' scenarios). Rob May expressed a similar concept in one of his last newsletters (that I highly recommend—you should definitely subscribe);
- **Reduce Catastrophic Risks Scenario**: an AI coded in a DAO with specific smart contracts will be able to only perform those actions, and nothing more (it will have a limited action space then).

In spite of all the benefits that AI will receive from an interaction with blockchain technologies, I do have one big question with no answer whatsoever.

AI was born as in an open-source environment where data was the real moat. With this data democratization (and open-source software) how can we be sure that AI will prosper and will keep being developed? What would be the new moat? My only guess at the moment? Talent…

4.5 Decentralized Intelligent Companies

There are plenty of landscapes of blockchain and cryptocurrencies startups out there. I am anyway only interested in those companies working at the intersection (or the *convergence*, as someone calls it) of AI and blockchain, which apparently are not that many. They are mainly concentrated in San Francisco area and London, but there are examples in New York, Australia, China, as well as some European countries.

They are indeed so few of them that is quite hard to classify them into clusters. I usually like to try to understand the underlying patterns and the type of impact/application certain groups of companies are having in the industry, but in this case is extremely difficult given the low number of data points so I will simply categorize them as follows:

- **Decentralized Intelligence**: TraneAI (training AI in a decentralized way), Neureal (peer-to-peer AI supercomputing), SingularityNET (AI marketplace),

Neuromation (synthetic datasets generation and algorithm training platform), AI Blockchain (multi-application intelligence), BurstIQ (healthcare data marketplace), AtMatrix (decentralized bots), OpenMinedproject (data marketplace to train machine learning locally), Synapse.ai(data and AI marketplace), Dopamine.ai (B2B AI monetization platform), and Effect.ai (decentralized AI workforce and services marketplace);
- **Conversational Platform**: Green Running (home energy virtual assistant), Talla (chatbot), and doc.ai (quantified biology and healthcare insights);
- **Prediction Platform**: Augur (collective intelligence) and Sharpe Capital(crowd-source sentiment predictions);
- **Intellectual Property**: Loci.io (IP discovery and mining);
- **Data Provenance**: KapeIQ (fraud detection on healthcare entities), Data Quarka (facts checking), Priops (data compliance), and Signzy (KYC);
- **Trading**: Euklid (bitcoin investments) and EthVentures (investments on digital tokens). For other (theoretical) applications in finance, see Lipton (2017);
- **Insurance**: Mutual.life (P2P insurance) and Inari (general);
- **Miscellaneous**: Social Coin (citizens' reward systems), HealthyTail (pet analytics), Crowdz (e-commerce), DeepSee (media platform), and ChainMind (cybersecurity).

A few general comments:

- It looks interesting that many AI–blockchain companies have **larger advisory board than teams**. It might be an early sign that the convergence is not fully realized yet and there are more things we don't understand that those ones we know;
- I am personally very excited to see the development of the first category (**decentralized intelligence**) but I also see a huge development in **conversational** and **prediction platforms** as well as **intellectual property**. I grouped other examples under "miscellaneous" because I don't think at this stage they represent a specific category but rather only single attempt to match AI and blockchain;
- **Those Companies are Incredibly Hard to Evaluate**. The websites are often cryptic enough to not really understand what they do and how (a bit paradoxical if you want to buy the blockchain transparency paradigm) and technology requires a high tech-education to be fully assessed. Cutting through the hype is a demanding task and this makes it very easy to be fooled. But let me give you a concrete example: **ever heard of Magos AI**? In the effort of researching companies for this post, I found myself reading several articles on this forecasting blockchain AI-driven platform company (Wired, PR Newswire, etc.), which just did an ICO for over half a million dollars and that made great promises on its deliverables. The website didn't work—weird, if you consider that they need to share material/information on the ICOs. But you know, it might happen. I made then an extra effort because I read it on Wired and I was curious to know more about it. I was able to find its co-founders, which I couldn't find eventually on Linkedin. Weird again. Well, there are people who do not like socials, fair enough, especially if you consider that until 3 months ago there was no proof of the company existence whatsoever. Let

> me look into the rest of the team. Nothing even there, and no traceable indications of their previous experiences (except for the CTO master in analytics, that I found no proof of). I tried to then dig into the technology: white papers, blue papers, yellow papers, you name it. I only found reviews of them, no original copies. Final two steps: I don't consider myself an expert in blockchain at all, but I read, a lot. And I also believe I am fairly knowledgeable when it comes to AI and what is happening in the industry. These guys claimed they created five different neural nets that could achieve the same accuracy in complex different domains than Libratus (or DeepStack) reached in Poker, but I never heard of them—very weird. Well, you know what? Maybe I could write them and meet them to understand. Their address points to the AXA office in Zurich. Ah. After 5 min of research, I finally Google the two keywords: "Magos scam". It seems these guys took the money and disappeared. They are surely building the six neural net somewhere, so stay tuned.

My point here is that exponential technologies are fantastic and can advance mankind, but as much as the benefits increase also the potential "*negative convergence*" increases exponentially. Stay alert.

4.6 Conclusion

Blockchain and AI are the two extreme sides of the technology spectrum: one fostering centralized intelligence on close data platforms, and the other promoting decentralized applications in an open-data environment. However, if we find an intelligent way to make them working together, the total positive externalities could be amplified in a blink.

There are of course technical and ethical implications arising from the interaction between these two powerful technologies, as for example how do we edit (or even forget) data on a blockchain? Is an ***editable blockchain*** the solution? And is not an AI–blockchain pushing us to become data hoarder?

Honestly, I think the only thing we can do is keep experimenting.

References

Castro, M., & Liskov, B. (2002). Practical Byzantine fault tolerance and proactive recovery. *ACM Transactions on Computer Systems, 20*(4), 398–461.

Catalini, C., & Gans, J. S. (2017). *Some simple economics of the blockchain*. MIT Sloan School Working Paper, pp. 5191–5116.

Deloitte. (2016). *Blockchain Enigma. Paradox. Opportunity*. White Paper.

Iansiti, M., & Lakhani, K. R. (2017, January–February). The truth about blockchain. *Harvard Business Review*, 118–127.

Lipton, A. (2017). Blockchains and distributed ledgers in retrospective and perspective. arXiv:1703.01505.

Mazieres, D. (2016). *The stellar consensus protocol: A federated model for internet-level consensus*. White Paper.

Mingxiao, D., Xiaofeng, M., Zhe, Z., Xiangwei, W., & Qijun, C. (2017, October 5–8). A review on consensus algorithm of blockchain. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. Canada: Banff Center, Banff.

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. White Paper.

O'Dwyer, K. J., & Malone, D. (2014). Bitcoin mining and its energy footprint. In *25th IET Irish Signals & Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies (ISSC 2014/CIICT 2014)*, Limerick (pp. 280–285).

Outlier Ventures. (2017). *Blockchain-enabled convergence*. White Paper.

Sasson, E. B., Chiesa, A., Garman, C., Green, M., Miers, I., Tromer, E., et al. (2014). Zerocash: Decentralized anonymous payments from bitcoin. In *2014 IEEE Symposium on Security and Privacy (SP)* (pp. 459–474).

Unicredit. (2016). *Blockchain technology and applications from a financial perspective*. Technical Report.

Vasin, P. (2014). *BlackCoin's proof-of-stake protocol v2*. White Paper.

Zyskind, G., Nathan, O., & Pentland, A. (2015). Enigma: Decentralized computation platform with guaranteed privacy. arXiv:1506.03471.

Chapter 5
The New CxO Gang: Data, AI, and Robotics

Abstract This chapter explains and identifies the emergence of new relevant figures in the data ecosystem space, namely the chief data officer, the chief AI officer, and the chief robotics officer. It will show the differences between them and highlight where they are needed and how they can be used efficiently.

5.1 Hiring New Figures to Lead the Data Revolution

It has been said that this new wave of exponential technologies will threaten a lot of jobs, both blue- and white-collar ones. But if on one hand, many roles will disappear, on the other hand in the very short-term, we are observing new people coming out from the crowd to lead this revolution and set the pace.

These are the people who really understand both the technicalities of the problems as well as have a clear view of the business implications of the new technologies and can easily plan how to embed those new capabilities in enterprise contexts.

Hence, I am going to briefly present three of them, i.e., the **Chief Data Officer** (CDO), the **Chief Artificial Intelligence Officer** (CAIO), and the **Chief Robotics Officer** (CRO). Sad to be said, I never heard about a "*Chief of Data Science*", but for some strange reasons, the role is usually called either "*Head of Data Science*" or "*Chief Analytics Officer*" (as if data scientist won't deserve someone at C-level to lead their efforts).

Let's see then who they are and what they would be useful for.

F. Corea, *Applied Artificial Intelligence: Where AI Can Be Used In Business*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-77252-3_5

5.2 The Chief Data Officer (CDO)

Apparently, it is a new role born in a lighter form straight after the financial crisis springing from the need to have a central figure to deal with technology, regulation, and reporting.

Therefore, the CDO is basically the guy who acts as a **liaison between the CTO (tech guy) and the CAO/Head of Data Science (data guy)** and takes care of data quality and data management.

Actually, her final goal is **to guarantee that everyone can get access to the right data in virtually no time**.

In that sense, a CDO is the guy in charge of "**democratizing data**" within the company.

It is not a static role, and it evolved from simply being a *facilitator* to being a *data governor*, with the tasks of defining data management policies and business priorities, shaping not only the data strategy, but also the frameworks, procedures, and tools. In other words, he is a kind of "Chief of Data Engineers" (if we agree on the distinctions between data scientists, who actually deal with modeling, and data engineers, who deal with data preparation and data flow).

> The difference between a CIO and CDO (apart from the words data and information…) is best described using the bucket and water analogy. The CIO is responsible for the bucket, ensuring that it is complete without any holes in it, the bucket is the right size with just a little bit of spare room but not too much and its all in a safe place. The CDO is responsible for the liquid you put in the bucket, ensuring that it is the right liquid, the right amount and that's not contaminated. The CDO is also responsible for what happens to the liquid, and making the clean vital liquid is available for the business to slake its thirst. (Caroline Carruthers, Chief Data Officer Network Rail, and Peter Jackson, Head of Data Southern Water)

Interestingly enough, the role of the CDO as we described it is both *vertical* and *horizontal*. It spans indeed across the entire organization even though the CDO still needs to report to someone else in the organizational chart. Who the CDO reports to will be largely determined by the organization he is operating in. Furthermore, it is also relevant to highlight that a CDO can be found more likely in larger organizations rather than small startups. The latter type is indeed usually set up to be data-driven (with a *forward-looking approach*) and, therefore, the CDO function is already embedded in the role who designs the technological infrastructure/data pipeline.

It is also true that not every company has a CDO, so how do you decide to eventually get one? Well, simply out of internal necessity, strict incoming regulation, and because all your business intelligence projects are failing because of data issues. If you have any of these problems, you might need someone who pushes the "fail-fast" principle as the data approach to be adopted throughout the entire organization, who considers data as a company asset and wants to set the fundamentals to allow fast trial and error experimentations. And above all, someone **who is centrally liable and accountable for anything about data**.

A CDO is then the **end-to-end data workflow responsible and it oversees the entire data value chain**.

Finally, if the CDO will do his job in a proper way, you'll be able to see two different outcomes: first of all, the board will stop asking for quality data and will have clear in mind what every team is doing. Second, and most important, a **good CDO aims to create an organization where a CDO has no reasons to exist**.

It is counterintuitive, but basically, a CDO will do a great job when the company won't need a CDO anymore because **every line of business will be responsible and liable for their own data**.

In order to reach his final goal, he needs to prove from the beginning that not investing in higher data quality and frictionless data transfer might be a source of inefficiency in business operations, resulting in non-optimized IT operations and making compliance as well as analytics much less effective.

5.3 The Chief Artificial Intelligence Officer (CAIO)

If the CDO is somehow an already consolidated role, the CAIO is nothing more than a mere industry hypothesis (not sure I have seen one yet, although the strong ongoing discussions between AI experts and sector players—see here and here for two opposite views on the topic). Moreover, the creation of this new role highlights the emergence of two different schools of thought of enterprise AI, i.e., **centralized versus decentralized AI implementation**, and a clear cost–benefit analysis to understand which approach that will work better is still missing.

My two cents are that elevating AI to be represented at the board level means to really become an AI-driven company and embed AI into every product and process within your organization—and I bet not everyone is ready for that.

So, let's try to sketch at a glance the most common themes to consider when talking about a CAIO:

- **Responsibilities** (*what he does*): a CAIO is someone who should be able to connect the dots and **apply AI across data and functional silos** (this is Andrew Ng's view, by the way). If you also want to have a deeper look at what a CAIO job description would look like, check out here the article by Tarun Gangwani;
- **Relevance** (*should you hire a CAIO?*): you only need to do it if you understand that **AI is no longer a competitive advantage to your business but rather a part of your core product** and business processes;
- **Skills** (*how do you pick the right guy?*): first and more important, a CAIO has to be a "guiding light" within the AI community, because he will be one of your decisive assets to **win the AI talent war**. This means that he needs to be highly respected and trusted, which is something that comes only with a strong understanding of **foundational technologies and data infrastructure**. Finally, being a cross-function activity, he needs to have the right balance between **willingness to risk and experiment to foster innovation** and **attention to product and company needs** (he needs to support different lines of business);

- **Risks** (*is a smart move hiring a CAIO?*): there are two main risks, which are (i) the misalignment between technology and business focus (you tend to put more attention on technology rather than business needs), and (ii) every problem will be tackled with AI tools, which might not be that efficient (this type of guys are super trained and will be highly paid, so it is natural they will try to apply AI to everything).

Where do I stand on that? Well, my view is that a CAIO is something which makes sense, **even though only temporarily**. It is an essential position to allow a smooth transition for companies who strive for becoming AI-driven firms, but I don't see the role to be any different from what a smart tech CEO of the future should do (of course, supported by the right lower management team). However, for the next decade having a centralized function with the task of **using AI to support the business lines** (50% of the time) and **foster innovation internally** (50% of the time) it sounds extremely appealing to me.

In spite of all the predictions I can make, the reality is that the relevance of a CAIO will be determined by how we will end up approaching AI, i.e., whether it will be eventually considered a mere instrument (**AI-as-a-tool**) or rather a proper business unit (**AI-as-a-function**).

5.4 The Chief Robotics Officer (CRO)

We moved from the CDO role, which has been around for a few years now, to the CAIO one, which is close to being embedded in organizational charts. But the Chief Robotics Officer is a completely different story.

Even if someone is speaking about the importance of it, it is really not clear what his tasks would be and what kind of benefits would bring to a company, and envisaging this role requires a huge leap of imagination and optimism about the future of work (and business).

In few words, what a CRO will be supposed to take care of is **managing the automated workforce of the company**. To use Gartner's words, "*he will oversee the blending of human and robotic workers.*" He will be responsible for the overall automatization of workflows and to integrate them smoothly into the normal design process and daily activities.

I am not sure I get the importance of this holistic approach to enterprise automation, although I recognize the relevance of having a central figure, who will actively keep track and communicate to employees all the changes made in transforming a manual activity/process into an automated one.

Another interesting point is who the CRO will report to, which is, of course, shaped by his real functions and goals. If robotics is deeply rooted into the company and allows to create or access new markets, **a CRO might directly report to the CEO**. If his goal is instead to automatize internal processes to achieve a higher efficiency,

he will likely report to the COO or to a strategic CxO (varying on industry and vertical).

My hypothesis is that this is going to be a **strategic role** (and not a technical one, as you might infer from the name) which, as the CAIO, might have a positive impact in the short term (**especially in managing the costs of adopting early robotics technologies**), but no reason to exist in the longer term. It is easier to think about it in physical product industries rather than digital products or services companies, but automation will likely happen in a faster way in the latter, so we will end up having a *Chief of Physical Robotics Officer* (to manage the supply chain workflow) as well as a *Chief of Digital Robotics Officer* (to manage instead the automation of processes and activities).

Chapter 6
Machine Ethics and Artificial Moral Agents

Abstract This final chapter concerns issues that are usually not associated with the business performance but that instead have a profound impact on the company's financial results. Ethics is indeed a strong component in the algorithmic development and should be managed with care. The chapter will discuss the most common ethics problems and data biases and propose some food for thoughts rather than solutions. It will also talk about the control problem, the accounting and explainability issues, and the development of a safe AI.

6.1 How to Design Machines with Ethically Significant Behaviors

There has been a lot of talk over the past months about AI being our best or worst invention ever. The chance of robots taking over and the following catastrophic sci-fi scenario makes the ethical and purposeful design of machines and algorithms not simply important but necessary.

But the problems do not end here. Incorporating ethical principles into our technology development process should not just be a way to prevent human race extinction but also a way to understand how to use the power coming from that technology responsibly.

This chapter does not want to be a guide for ethics for AI or setting the guidelines for building ethical technologies. It is simply a stream of consciousness on questions and problems I have been thinking and asking myself, and hopefully, it will stimulate some discussion.

Now, let's go down the rabbit hole…

F. Corea, *Applied Artificial Intelligence: Where AI Can Be Used In Business*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-77252-3_6

6.2 Data and Biases

The first problem everyone raises when speaking about ethics in AI is, of course, about data. Most of the data we produce (if we exclude the ones coming from observation of natural phenomena) are artificial creations of our minds and actions (e.g., stock prices, smartphone activity, etc.). As such, **data inherit the same biases we have as humans**.

First of all, what is a cognitive bias? The (maybe controversial) way I look at it is that a cognitive bias is a **shortcut of our brain that translates into behaviors, which required less energy and thought to be implemented**. So, a bias is a good thing to me, at least in principle. The reason why it becomes a bad thing is that the external environment and our internal capacity to think do not proceed pari passu. Our brain gets trapped into heuristics and shortcuts which could have resulted into competitive advantages 100 years ago, but is not that plastic to quickly adapt to the change of the external environment (I am not talking about a single brain, but rather on a species level).

In other words, *the systematic deviation from a standard of rationality or good judgment* (this is how bias is defined in psychology) is nothing more for me than a simple **evolutionary lag of our brain**.

Why all this *excursus*? Well, because I think that most of the biases data embed comes from our own cognitive biases (at least for data resulting from human and not natural activities). There is, of course, another block of biases which stems from pure statistical reasons (*the expected value is different from the true underlying estimated parameter*). Kris Hammond of Narrative Science merged those two views and identified at least five different biases in AI. In his words:

- **Data-driven bias** (bias that depends on the input data used);
- **Bias through interaction**;
- **Similarity bias** (it is simply the product of systems doing what they were designed to do);
- **Conflicting goals bias** (systems designed for very specific business purposes end up having biases that are real but completely unforeseen);
- **Emergent bias** (decisions made by systems aimed at personalization will end up creating bias "bubbles" around us).

But let's go back to the problem. How would you solve the biased data issue then?

Simple solution: you can try to remove any data that could bias your engine ex ante. Great solution, it will require some effort at the beginning, but it might be feasible.

However, let's look at the problem from a different angle. I was educated as an economist, so allow me to start my argument with this statement: **let's assume we have the perfect dataset**. It is not only omni-comprehensive but also clean, consistent and deep both longitudinally and temporally speaking.

Even in this case, **we have no guarantee AI won't learn the same bias autonomously as we did**. In other words, removing biases by hand or by construction is not a guarantee of those biases do not come out again spontaneously.

This possibility also raises another (philosophical) question: we are building this argument from the assumption that *biases are bad* (mostly). So let's say the machines come up with a result we see as biased, and therefore we reset them and start again the analysis with new data. But the machines come up with a similarly "biased result". Would we then be open to accepting that as true and revision what we consider to be biased?

This is basically a **cultural and philosophical clash between two different species**.

In other words, I believe that two of the reasons why embedding ethics into machine designing is extremely hard is that (i) **we don't really know unanimously what ethics is**, and (ii) we should be open to admit that our values or ethics might not be completely right and that what we consider to be biased is not the exception but rather the norm.

Developing a (general) AI is making us think about those problems and **it will change** (if it hasn't already started) **our values system**. And perhaps, who knows, we will end up learning something from *machines*' *ethics* as well.

6.3 Accountability and Trust

Well, now you might think the previous one is a purely philosophical issue and that you probably shouldn't care about it. But the other side of the matter is about how much you **trust your algorithms**. Let me give you a different perspective to practically looking at this problem.

Let's assume you are a medical doctor and you use one of the many algorithms out there to help you diagnose a specific disease or to assist you in a patient treatment. In the 99.99% of the time the computer gets it right—and it never gets tired, it analyzed billions of records, it sees patterns that a human eye can't perceive, we all know this story, right? But what if in the remaining 0.01% of the case your instinct tells you something opposite to the machine result and you end up to be right? What if you decide to follow the advice the machine spit out instead of yours and the patient dies? Who is liable in this case?

But even worse: let's say in that case you follow your gut feeling (we know is not gut feeling though, but simply your ability to recognize at a glance something you know to be the right disease or treatment) and you save a patient. The following time (and patient), you have another conflict with the machine results but strong of the recent past experience (because of a *hot-hand fallacy* or an *overconfidence bias*) you think to be right again and decide to disregard what the artificial engine tells you. Then the patient dies. Who is liable now?

The question is quite delicate indeed and the scenarios in my head are as follows:

(a) a scenario where the doctor is only human with no machine assistance. The payoff here is that liability stay with him, he gets it right 70% of the time, but the things are quite clear and sometimes he gets right something extremely hard (the lucky guy out of 10,000 patients);

(b) a scenario where a machine decides and gets it right 99.99% of the time. The negative side of it is an unfortunate patient out of 10,000 is going to die because of a machine error and the liability is not assigned to either the machine or the human;

(c) a scenario the doctor is assisted but has the final call to decide whether to follow the advice. The payoff here is completely randomized and not clear to me at all.

As a former economist, I have been trained to be heartless and reason in terms of expected values and big numbers (basically a **Utilitarian**), therefore scenario b) looks the only possible to me because it saves the greatest number of people. But we all know is not that simple (and of course doesn't feel right for the unlucky guy of our example): think about the case, for instance, of autonomous vehicles that lose controls and need to decide if killing the driver or five random pedestrians (the famous ***Trolley Problem***). Based on that principles I'd save the pedestrians, right? But what about all those five are criminals and the driver is a pregnant woman? Does your judgment change in that case? And again, what if the vehicle could instantly use cameras and visual sensors to recognize pedestrians' faces, connect to a central database and match them with health records finding out that they all have some type of terminal disease? You see, the line is blurring…

The final doubt that remains is then not simply about liability (and the choice between pure outcomes and ways to achieve them) but rather on trusting the algorithm (and I know that for someone who studied 12 years to become doctor might not be that easy to give that up). In fact, **algorithm adversion** is becoming a real problem for algorithms-assisted tasks and it looks that people want to have an (even if incredibly small) degree of control over algorithms (Dietvorst et al. 2015, 2016).

But above all: **are we allowed to deviate from the advice we get from accurate algorithms**? And if so, in what circumstances and to what extent?

If an AI would decide on the matter, it will also probably go for scenario b) but we as humans would like to find a compromise between those scenarios because we "*ethically*" don't feel any of those to be right. We can rephrase then this issue under the "**alignment problem**" lens, which means that the goals and behaviors an AI have need to be aligned with human values—an AI needs to think as a human in certain cases (but of course the question here is how do you discriminate? And what's the advantage of having an AI then? Let's, therefore, simply stick to the traditional human activities).

In this situation, the work done by the Future of Life Institute with **the Asilomar Principles** becomes extremely relevant.

The alignment problem, in fact, also known as "**King Midas problem**", arises from the idea that no matter how we tune our algorithms to achieve a specific objective, we are not able to specify and frame those objectives well enough to prevent the machines to pursue undesirable ways to reach them. Of course, a theoretically viable solution would be to let the machine maximizing for our true objective without setting it ex ante, making, therefore, the algorithm itself free to observe us and understand what we really want (as a species and not as individuals, which might entail also the possibility of switching itself off if needed).

Sounds too good to be true? Well, maybe it is. I indeed totally agree with Nicholas Davis and Thomas Philbeck from WEF that in the Global Risks Report 2017 wrote:

> There are complications: humans are irrational, inconsistent, weak-willed, computationally limited and heterogeneous, all of which conspire to make learning about human values from human behavior a difficult (and perhaps not totally desirable) enterprise.

What the previous section implicitly suggested is that not all AI applications are the same and that *error rates* apply differently to different industries. Under this assumption, it might be hard to draw a line and design an accountability framework that does not penalize applications with weak impact (e.g., a recommendation engine) and at the same time do not underestimate the impact of other applications (e.g., healthcare or AVs).

We might end up then designing **multiple accountability frameworks** to justify algorithmic decision-making and mitigate negative biases.

Certainly, the most straightforward solution to understand who owns the liability for a certain AI tool is thinking about the following threefold classification:

- ***We should hold the AI system itself as responsible for any misbehavior*** (does it make any sense?);
- ***We should hold the designers of the AI as responsible for the malfunctioning and bad outcome*** (but it might be hard because usually AI teams might count hundred of people and this preventative measure could discourage many from entering the field);
- ***We should hold accountable the organization running the system*** (to me it sounds the most reasonable between the three options, but I am not sure about the implications of it. And then what company should be liable in the AI value chain? The final provider? The company who built the system in the first place? The consulting business which recommended it?).

There is not an easy answer and much more is required to tackle this issue, but I believe a good starting point has been provided by Sorelle Friedler and Nicholas Diakopoulos. They suggest to consider accountability through the lens of five core principles:

- **Responsibility**: a person should be identified to deal with unexpected outcomes, not in terms of legal responsibility but rather as a single point of contact;
- **Explainability**: a decision process should be explainable not technically but rather in an accessible form to anyone;
- **Accuracy**: *garbage in, garbage out* is likely to be the most common reason for the lack of accuracy in a model. The data and error sources need then to be identified, logged, and benchmarked;
- **Auditability**: third parties should be able to probe and review the behavior of an algorithm;
- **Fairness**: algorithms should be evaluated for discriminatory effects.

6.4 AI Usage and the Control Problem

Everything we discussed so far was based on two implicit assumptions that we did not consider up to now: first, everyone is going to benefit from AI and everyone will be able and in the position to use it.

This might not be completely true though. Many of us will indirectly benefit from AI applications (e.g., in medicine, manufacturing, etc.) but we might live in the future in a world where only a handful of big companies drives the AI supply and offers fully functional AI services, which might not be affordable for everyone and above all, not *super partes*.

AI democratization versus a centralized AI is a policy concern that we need to sort out today: if from one hand the former increases both the benefits and the rate of development but comes with all the risks associated with system collapse as well as malicious usages, the latter might be more safe but unbiased as well. Should AI be centralized or for everyone?

The second hypothesis, instead, is that we will be forced to use AI with no choice whatsoever. This is not a light problem and we would need a higher degree of education on what AI is and can do for us to not be misled by other humans. If you remember the healthcare example we described earlier, this could be also a way to partially solve some problem in the accountability sphere. If the algorithm and the doctor have a contradictory opinion, you should be able to choose who to trust (and accepting the consequences of that choice).

The two hypothesis above described lead us to another problem in the AI domain, which is the ***Control Problem***: if it is centralized, who will control an AI? And if not, how should it be regulated?

I wouldn't be comfortable at all to empower any government or existing public entity with such a power. I might be slightly more favorable to a big tech company, but even this solution comes with more problems than advantages. We might then need a new impartial organization to decide how and when using an AI, but history teaches us we are not that good in forming mega impartial institutional players, especially when the stake is so high.

Regarding the AI decentralization instead, the regulation should be strict enough to deal with cases such as **AI-to-AI conflicts** (what happens when 2 AIs made by two different players conflict and give different outcomes?) or the ethical use of a certain tool (a few companies are starting their own **AI ethics board**) but not so strict to prevent research and development or full access to everyone.

I will conclude this section with a final question: I strongly believe there should be a sort of "***red button***" to switch off our algorithms if we realize we cannot control it anymore. However, the question is who would you grant this power to?

6.5 AI Safety and Catastrophic Risks

As soon as AI will become a commodity, it will be used maliciously as well. This is a virtual certainty. And the value alignment problem showed us that we might get in trouble due to a variety of different reasons: it might be because of misuses (**misuse risks**), because of some accident (**accident risks**), or it could be due to **other risks**.

But above all, no matter the risk we face, it looks that AI is dominated by some sort of exponential chaotic underlying structure and getting wrong even minor things could turn into catastrophic consequences. This is why is paramount to understand every minor nuance and solve them all without underestimating any potential risk.

Amodei et al. (2016) actually dug more into that and drafted a set of five different core problems in AI safety:

- **Avoiding negative side effects**;
- **Avoiding reward hacking**;
- **Scalable oversight** (respecting aspects of the objective that are too expensive to be frequently evaluated during training);
- **Safe exploration** (learning new strategies in a non-risky way);
- **Robustness to distributional shift** (can the machine adapt itself to different environments?).

This is a good categorization of AI risks but I'd like to add the ***interaction risk*** as fundamental as well, i.e., the way in which we interact with the machines. This relationship could be beneficial (see the ***Paradigm 37–78***) but comes with several risks as well, as for instance the so-called ***dependence threat***, which is a highly visceral dependence of human on smart machines.

A final food for thought: we are all advocating for full transparency of methods, data, and algorithms used in the decision-making process. I would also invite you though to think that full transparency comes with the great **risk of higher manipulation**. I am not simply referring to cyber attacks or bad-intentioned activities, but more generally to the idea that once the rules of the game are clear and the processes reproducible, it is easier for anyone to hack the game itself.

Maybe companies will have specific departments in charge of influencing their own or their competitors' algorithms, or there will exist companies with the only scope of altering data and final results. Just think about that…

6.6 Bonus Paragraph: 20 Research Groups on AI Ethics and Safety

There are plenty of research groups and initiatives both in academia and in the industry start thinking about the relevance of ethics and safety in AI. The most known ones are the following 20, in case you like to have a look at what they are doing:

- Future of Life Institute (Boston);
- Berkman Klein Center (Boston);
- Institute Electrical and Electronic Engineers—IEEE (Global);
- Centre for the Study on Existential Risks (Cambridge, UK);
- K&L Gates Endowment for Ethics (Global);
- Center for Human-Compatible AI (Berkeley, CA);
- Machine Intelligence Research Institute (Berkeley, CA);
- USC Center for Artificial Intelligence in Society (Los Angeles);
- Leverhulme Center for the Future of Intelligence (Cambridge, UK);
- Partnership on AI (Global);
- Future of Humanity Institute (Oxford, UK);
- AI Austin (Austin, US);
- Open AI (San Francisco, US);
- Campaign to Stop Killer Robots (Global);
- Campaign Against Sex Robots (Global);
- Foundation for Responsible Robotics (Global);
- Data & Society (New York, USA);
- World Economic Forum's Council on the Future of AI and Robotics(Global);
- AI Now Initiative (New York, USA);
- AI 100 (Stanford, CA).

Finally, Google has just announced the "People+AI research" (PAIR) initiative, which aims to advance the research and design of people-centric AI systems.

6.7 Conclusion

Absurd as it might seem, I believe ethics is a technical problem. Writing this post, I realized how much little I know and even understand about those topics. It is incredibly hard to have a clear view and approach to ethics in general, let's not even think about the intersection of AI and technology. I didn't even touch upon other questions that should keep AI experts up at night (e.g., unemployment, security, inequality, universal basic income, robot rights, social implications, etc.) but I will do in future posts (any feedback would be appreciated in the meantime).

I hope your brain is melting down as mine in this moment, but I hope some of the above arguments stimulated some thinking or ideas regarding new solutions to old problems.

I am not concerned about robots taking over or Skynet terminates us all, but rather of humans using improperly technologies and tools they don't understand. I think that the sooner we clear up our mind around those subjects, the better it would be.

References

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv:1606.06565v2.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology, 144*(1), 114–126.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. Available at SSRN: https://ssrn.com/abstract=2616787 or http://dx.doi.org/10.2139/ssrn.2616787.